To: Laidlaw, Tina[Laidlaw.Tina@epa.gov]
Cc: Urban, Eric[EUrban@mt.gov]; Watson, Vicki (vicki.watson@mso.umt.edu)[vicki.watson@mso.umt.edu]
From: Suplee, Mike
Sent: Tue 4/30/2013 4:35:20 PM
Subject: Response to comments, wadeable stream criteria
Memo_Update1_ResponsetoReviews_FNL.pdf

Hi Tina;

Attached is our final response to comments from the peer reviewers on the wadeable stream criteria document. If you recall, last August three anonymous peer reviewers reviewed the document via the NSTEPS criteria process.

The final technical document (Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers: Update 1) is going through final formatting and should be posted on the department's internet site in a week or so. The changes to the document discussed in the attached memo will be reflected in the final technical report.

Thanks, and let me know if you have any questions.

Mike

Montana Department of
ENVIRONMENTAL QUALITY

MEMO

1520 E 6th Avenue
PO Box 200901
Helena MT 59620-0901
(406) 444-6697 FAX: (406) 444-3836

To: Tina Laidlaw, U.S Environmental Protection Agency

CC: Eric Urban, Head, Water Quality Standards Section

From: Michael Suplee, Ph.D., Water Quality Standards Section; Vicki Watson, Ph.D., University of Montana

Date: 4/30/2013

RE: MT DEQ's response to peer-review comments on "Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers: Addendum 1"

Reviews were received by three anonymous peer reviewers on the document referenced above in August, 2012. The reviewers were selected by the U.S. EPA in conjunction with the Nutrient Scientific Technical Exchange Partnership & Support (NSTEPS) service. One of the services NSTEPS provides is review of state-developed numeric nutrient criteria.

Section 1.0 below addresses comments that were common to two or all three reviewers; MT DEQ's response is provided in each case. **Section 2.0** lists salient comments from individual reviewers. **Section 3.0** summarizes changes to the "Addendum 1" document[1] that will be made as a result of reviewers' comments. Some comments were minor or editorial in nature and these have simply been addressed during the finalization of the document.

1.0 Comments from Peer Reviewers Addressing EPA's Six Core Questions

EPA posed six questions to the reviewers. The first queried their overall impression of the approach MT DEQ took to derive the numeric nutrient criteria. There was universal agreement among the three reviewers that the approach taken was thorough, scientifically sound, and an effective use of available and relevant information. There were, of course, concerns and recommendations as well. The five remaining EPA questions are addressed in each of the sections below (**Section 1.1** to **Section 1.5**) and most of the reviewer's comments/concerns are covered in these sections. **Sections 1.6 and 1.7** address other issues raised by the peer reviewers.

[1] The draft document was called "Addendum 1" because we considered it an extension of methods and ideas put forth in Suplee et al. (2008). However, enough material has changed and the document is now sufficiently stand-alone that in final form it has been named "Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers: Update 1".

1.1 Concern the MT DEQ has not Provided Nutrient Criteria Recommendations for the Level IV Ecoregion "River Breaks"

Two reviewers were concerned that MT DEQ did not provide draft criteria for this ecoregion. They wanted to see more reference sites, recommended that MT DEQ discuss soluble nutrients from the ecoregion's reference sites, and discuss the potential impact on downstream uses if no criteria were adopted in this area. A third reviewer was apparently very familiar with western plains environments, and understood our reasoning, but was still concerned about downstream use impacts.

RESPONSE: The basic tenant of MT DEQ's approach is to apply appropriate stressor-response studies to a region and then compare the harm-to-use thresholds derived from the studies to the reference distribution (Suplee et al., 2007). Although there are just eight reference sites in the River Breaks, it is not so small a dataset as to preclude reasonable comparisons to dose-response studies. Compiling the reference data by site medians (see discussion on this topic in **Section 1.6** below) did not substantially alter the plains region dose-response-to-reference matches (equal to the 48th percentile for the median dataset and the 53rd percentile for the all-observations dataset, for TP; 60th and the 53rd,respectively, for TN). These data show, regardless of how the reference data are summarized, that harm-to-use concentrations applicable to the plains align with nutrient concentrations which are about average in the River Breaks' reference streams (i.e., River Breaks streams are naturally eutrophied or may have already responded to global increases in nitrogen loading [Vitousek et al., 1997]).

Per reviewers' recommendations, we have included soluble nutrients in the final report for the River Breaks. In relation to stream algal growth, it results that soluble nutrients in the River Breaks reference streams are already saturated, or are nearly so. For NO_2+NO_3, the 75th percentile of River Breaks reference streams was 241 or 631 µg N/L (median or all-observations datasets, respectively). Rier and Stevenson (2006) show there is little peak algal biomass increase above 308 µg soluble nitrogen per liter (and peak biomass may actually be saturated closer to 250 µg DIN/L). As such, River Breaks reference sites are often saturated with soluble nitrogen, which is the nutrient most likely to be added to these streams if future development were to occur. Soluble P concentrations are also high. At the 75th percentile of reference SRP is 18 or 20 µg P/L (median vs. all-observations datasets, respectively) and therefore these low-gradient reference streams are often P saturated for peak algal biomass, or nearly so (Horner et al., 1983; Bothwell, 1989).

The absence of numeric nutrient standards in the River Breaks does not mean there will be no nutrient controls whatsoever applied to new permitted sources. Aquatic life ammonia standards still apply year-round. Median pH in the River Breaks reference streams is 8.8, and with typical summer temperatures of 20 to 25°C, the ammonia criterion would be about 360 µg NH_{3+4}-N/L (DEQ-7, 2012) and would provide protection from the toxic effects of ammonia on early fish life stages. If all this ammonia were oxidized to nitrate the resulting nitrate concentration would be well within the nitrate range observed in the River Breaks reference sites. The human health standards of 1.0 mg NO_2-N/L and 10 mg NO_3-N/L would apply year round as well. Thus, Montana's existing water quality standards would preclude the River Breaks from becoming an 'industrial dumping ground', a concern expressed by one reviewer.

Downstream uses will be addressed in permitting situations via application of nondegradation. The River Breaks ecoregion basically drains directly into the mainstem Missouri and Yellowstone rivers. In spite of the elevated nitrates and total N and P coming from the River Breaks, MT DEQ has not observed nutrient problems in the lower Yellowstone River, i.e., algal levels at unacceptable levels or DO and pH that violate state quality standards. Summertime concentrations in the Yellowstone River near Glendive (in the heart of River Breaks country) during low-flow years average 490 µg TN/L and 55 µg TP/L, and are well below our recommended numeric nutrient criteria for the lower Yellowstone River during low flow(815 µg TN/L and 95 µg TP/L; Flynn and Suplee, 2013). In establishing any permit which would allow an N or P discharge that is likely to reach the Yellowstone or Missouri River, nondegradation would be considered.

We conclude that there is no scientifically-defensible way to derive numeric nutrient criteria for the control of eutrophication for streams of the River Breaks. The streams are highly turbid, flashy, have low levels of benthic algae and macrophytes, and have soluble nutrient concentrations at levels that saturate algal growth much of the time. Other MT DEQ programs will address impacts to downstream uses. We will not be recommending nutrient criteria for these streams in the final report.

1.2 Peer Reviewers' Views Concerning the Allowable 20% Exceedence Rate Associated with the Criteria (Pertains to Assessment Methodology[2])

RESPONSE: This topic closely ties to the topic in **Section 1.5** below, and is addressed there.

1.3 Peer Reviewers' Comments on MT DEQ's Use of Benthic Chlorophyll *a*, How the Chlorophyll *a* Threshold (125 mg Chl*a*/m^2) was Derived, and Thoughts on Other Biological Measurements Used to Support Eutrophication Assessment

Two reviewers found the use of benthic chlorophyll *a* to be an excellent tool for assessing eutrophication, while the third did not like it. Derivation of the chlorophyll *a* thresholds were considered appropriate although two reviewers felt that the threshold of 125 mg Chl*a*/m^2 may be too close to the harm-to-use threshold. (The third reviewer found it acceptable.) One reviewer notes that macroinvertebrates are a poor indicator of eutrophication.

RESPONSE: MT DEQ has had good success with measuring benthic chlorophyll *a* and does not believe the concerns of one reviewer (it's too variable, affected by grazers) apply to the physiographic regions where it is used. Note also that MT DEQ collects benthic ash free dry mass (as g/m^2), which can provide good indication of heavy benthic algal growth even if chlorophyll *a* levels have declined due to senescence. As pointed out by one reviewer, MT DEQ has a long tradition of measuring benthic algae density and diatom taxa and both of these are main features of the nutrient assessment method. We agree that macroinvertebrates are not an ideal tool for pinpointing eutrophication problems, which is why they are used secondarily, i.e., only after the better tools (benthic chlorophyll *a*, diatom metrics) have already been played out. At a recent conference of academic experts on stream ecology (April 16-18, 2013, Washington, D.C.), which one of the authors was fortunate enough to attend, there was wide agreement that macroinvertebrates have generally poor predictive power for eutrophication assessment.

MT DEQ had extensive internal discussion about where to set the benthic algae density after the results from the dosing study (Appendix B, Suplee and Sada de Suplee, 2011) showed that average levels of 127

---

[2] MT DEQ's assessment methodology for assessing eutrophication in wadeable streams(Suplee and Sada de Suplee, 2011) was completely revised in 2009-2010, went through public comment (including EPA review), and was finalized prior to the time that it was provided to the peer reviewers here.

mg Chl$a$/m^2 could result in seasonal DO problems. To inform that discussion, a mechanistic model was built to simulate the DO impact observed in the dosing study and the model showed that higher gradient streams in western Montana would not develop low DO due to their reaeration; lower gradient streams, however, would be impacted. In streams with good re-aeration, therefore, harm-to-use would not occur until 150 mg Chla/m^2 (the recreational threshold). To avoid creating an overly-complex application of the algae threshold, involving not only ecoregions but different beneficial uses and different benthic algae levels for different stream gradients, it was decided that one algae threshold would be established (125 mg Chl$a$/m^2) that should be largely protective of both aquatic life and recreation. Monitoring staff with experience using the thresholds understood the rationale and indicated that they were comfortable with it because, in most cases, streams' algae densities are well below or well above the thresholds, precluding borderline decisions. MT DEQ is measuring diel DO concentrations much more frequently now and will continue to evaluate the 125 mg Chl$a$/m^2 threshold; it can be readjusted if needed in the future.

1.4 Peer Reviewers' Assessment of MT DEQ's Reach-specific Nutrient-criteria Derivation Method

All three reviewers supported the approach taken.

RESPONSE: We are delighted that all three peer reviewers were very supportive of the approach that was taken.

1.5 Peer Reviewers' Views Concerning the use of the Binomial Test, its 20% Allowable Exceedence Rate, and the Student's T-test (Pertains to Assessment Methodology)

The reviewers' main questions and thoughts/observations pertaining to these subjects are summarized as follows. Questions: (a) By using an allowable exceedence rate of 20% and an effect size of 15%, is the exact binomial testing whether 35% of observations must exceed the criterion to be considered non-compliant? (b) Is it appropriate to use an effect size in the T-test? Thoughts/observations: (a)The T-test is a parametric test with assumptions of a normal distribution which are not the norm for the datasets being evaluated, and so it will be less likely to detect a difference in the mean of a nutrient dataset relative to the criterion, and (b) for the T-test to establish non-compliance, the average concentration of a test stream would need to be substantially above the 75th percentile of reference (the reviewer's presumed level at which protection of uses is assured) and this is under protective.

In addition, there was general confusion among reviewers on how MT DEQ has defined an observation when assessing nutrients, how reaches are delineated, and how statistical tests and biological information all fit together in the final assessment.

RESPONSE: MT DEQ's statistical assessment of nutrient concentrations in a stream segment can be reduced to two simple ideas: (1) a test (exact binomial) to determine the proportion of samples that exceed the criterion, and (2) a test (one sample Student's T-test for the mean) to help to identify when the average nutrient concentration has been pulled above the criterion, which may result because most samples are above the criterion or because just a few high outliers are. Each test is discussed below, followed by an overall conclusion.

<u>Binomial Test</u>. Excellent empirical data were available to MT DEQ to derive the allowable exceedence rate used in the test (more on this in a moment). Besides the all-important exceedence rate, MT DEQ had to give consideration to other factors for statistical testing including the realistic number of independent samples that could be collected in a stream reach (restrained by cost/time), and the desire to balance type I and II error rates, i.e., give roughly equal weight to the importance of error of over-regulation vs. failure to protect the environment (Mapstone, 1995). Realistic sample sizes were about 10 to 15, and as such it was impossible to have alpha and beta error both around 0.05 (95% confidence) because sample size would then need to be around 75. So MT DEQ opted for less confidence, i.e., alpha and beta error rates both ≈0.25 (75% confidence).

Allowable exceedence rate (number of samples allowed above the criterion while assuring the river supports beneficial uses) was empirically derived from long-term work on the Clark Fork River—a river where adopted nutrient standards are virtually identical to those proposed for western Montana streams. **The Clark Fork River analysis shows that a defensible criteria exceedence rate could range from 5-31%.** Twenty percent was identified as the most reasonable value. <u>To date, MT DEQ has not found or been made aware of another dataset by which an allowable exceedence rate for numeric nutrient criteria could be determined</u>. Because of this, MT DEQ will continue to use the 20% exceedence rate.

MT DEQ uses a 15% effect size. By establishing 15% effect size, MT DEQ is saying that this is the range of true exceedence rates where the consequence of decision errors is relatively minor. As a point of comparison, if there were a pollutant for which the allowable exceedence rate is set at 10% and it is known that virtually no impact will occur at 9% exceedence, but terrible impacts occur at 11% exceedence, then the effect size would have to be set very close to zero, because the consequence of decision error is huge. And as a result, very large numbers of samples may need to be collected to discern with accuracy that fine a cut on the exceedence rate. But for nutrients, the state-of-the science is still limited and what we do know tells us there is a fairly wide range (5-31%) where decision error impacts are minor; MT DEQ addressed this by selecting a somewhat wide (15%) effect size.

In the binomial—with 20% allowable exceedence rate and 15% effect size—MT DEQ is establishing that streams with <5% exceedence will always PASS (be found compliant with) the binomial test, and streams with >35% exceedence will always FAIL (be found non-compliant with) the binomial. (The reviewer is correct that the 20% exceedence rate and 15% effect size are additive.) Streams falling in between will sometimes PASS, sometimes FAIL (depends on **n**). It could be reasonably argued that 35% exceedence is too high, but sample-size reality then enters the picture: if we lower the effect size to 10%, i.e. streams with 30% exceedence rate will always FAIL the binomial, we would have to collect 25 samples to roughly balance alpha and beta error; too many samples to institute for routine stream nutrient monitoring. Other combinations of exceedance rates and effect sizes within defensible ranges (and again balancing alpha, beta error) also led to **n**'s in the low 20s or higher. In the end, MT DEQ settled on the exceedence rate and effect size we are currently using. However, note that in borderline situations (i.e., the assessment decision is not clear) MT DEQ will collect more data, and may very well end up with sample sizes closer to 20.

<u>T-test.</u> Per the reviewer's question, no, effect size is <u>not</u> included in the T-test. MT DEQ believes that at this point the EPA-recommended T-test is satisfactory for its purpose within the assessment methodology. It is robust against moderate deviations from normality (and many of the small datasets that are considered are essentially normally distributed). The reviewer is correct that the T-test loses power when datasets are highly skewed (and some of the datasets are skewed). But in actual cases where there are a few very large outliers among the 12 or so samples (this is a common scenario), the T-test still FAILS (indicates non-compliance, as we would want it to) even if the exact test statistics (p value, etc.) may not be particularly accurate. Staff who routinely carry out eutrophication assessments have expressed that the T-test results are largely in alignment with the totality of information provided by the binomial and biological measurements.

Regarding the idea that the average concentration in a test site would have to be much greater than the 75^{th} of reference in order to FAIL the T-test, two points can be made. (1) The same reviewer stated that nutrient criteria are best if based upon dose-response studies. MT DEQ has found that dose-response studies often show concentrations >75^{th} of reference are protective of legally-defined beneficial uses (Suplee et al., 2007; Suplee et al., 2008). Thus, PASSING the T-test because the average concentration in a test site is >75^{th} percentile of reference is not necessarily under protective. (2) MT DEQ uses a different test hypothesis depending on the stream's 303(d) listing history for nutrients. Already-listed streams have the null as "stream is impaired" and the alternative as "not impaired". Thus, MT DEQ has the most control on alpha error which is defined upfront in the test. This approach is more protective.

MT DEQ Procedures and Assumptions. Regarding clear explanations of MT DEQ procedures, MT DEQ laid out the entire assessment approach and its assumptions in Suplee and Sada de Suplee (2011), including a number of examples that can be followed (see Section 3.2.4 of that document). However it appears that the final element of the method, the data-review matrix contained in the Excel spreadsheet "NtrntAssessFramework.xlsx", may not have been seen by some reviewers. Lacking this final piece would have led to confusion for sure. In any case, MT DEQ believes that "Addendum 1" (now Update 1) is not the place to detail assessment methodologies that are well covered in other documents. Going forward, Update 1 will continue to focus on nutrient criteria and their derivation.

Conclusion. The binomial test and the T-test in MT DEQ's assessment methodology will continue to be used as configured. As noted by a reviewer, MT DEQ compensates for the higher-than-ideal FAIL threshold of 35% in the binomial test by establishing different null hypothesis depending on if the stream is (or is not) already listed on the 303(d) list, by including the T-test, and by lowering the chlorophyll *a* threshold to 125 mg $Chla/m^2$ (instead of 150 mg $Chla/m^2$). As noted by another reviewer, "Some of the quibbling on these values may never be resolved (including mine), and Montana needs to use best judgment supported by its analysis and other scientific results." We couldn't agree more.

1.6 Number of Reference Sites, Manner by which MT DEQ Characterizes the Reference Condition

One reviewer felt there were too few reference data. Reviewers felt that MT DEQ's novel use of the Brillouin Evenness Index should be (at a minimum) clearly spelled out, and include the equations. One reviewer felt that the Brillouin method was "interesting", but that it did not directly address the issue of temporal pseudoreplication which may arise in repeated measurements of nutrients at reference sites. The reviewer recommended a more traditional approach to summarize reference data, whereby each reference site's nitrogen and phosphorus observations are reduced to a site median, and then distribution statistics on the population of medians is calculated.

RESPONSE: Regarding the number of reference sites, MT DEQ believes it has a good reference site network and has been actively identifying and sampling reference sites for the past twelve years. (Limited work was also carried twenty years ago by Bahls et al. [1992].) From 2000 to 2009 much effort went towards identifying new sites. In some parts of the state (e.g., eastern Montana) staff has gone over the landscape several times and we are at the point where few if any new sites can readily be identified. As it stands, there are 185 different reference sites across the state and all major ecoregions are represented. Because of the relatively large overall number of sites, MT DEQ management indicated that the Reference Project should focus on resampling the network rather than seeking new sites. Some level III ecoregions (e.g., Idaho Batholith) would benefit from additional sites but it is unlikely that will occur in the near future.

We agree that the Brillouin Evenness Index formula should be provided in Update 1 with an explanation of why this approach was taken. This has been included in the final report. Regarding our use of the Brillouin Evenness Index vs. site medians to summarize the reference data, we offer the following. By taking the Brillouin Evenness Index approach, MT DEQ made the assumption that each nutrient observation in the dataset was independent even if collected from the same site. The vast majority of sample observations from the reference sites were collected a month apart, and MT DEQ has shown that such samples are usually temporally independent (Appendix A.3, Suplee and Sada de Suplee, 2011). We believe the data, after application of the evenness index to assure equitable representation of each site, provide a very valuable characterization of reference condition especially when a reader wants to know the true range of nutrient observations (minimum, maximum) in Montana reference sites during baseflow.

Stakeholders from the Montana Nutrient Work Group had earlier indicated that this was important to them. And as pointed out by one reviewer, the exact manner by which reference data are summarized is not terribly important because we do not carry out inferential statistics with the data, nor are criteria tied to a specific reference percentile. We agree with the reviewer that with the approach we used we cannot assure that there is no intra-site temporal pseudoreplication, an issue discussed at length in Hurlbert (1984). In response, we have now provided two summary statistics tables for each ecoregion; the original (derived, as before, using all observations and the Brillouin Evenness Index), and a 2nd table which shows the frequency distribution (25th through 90th) based on the median nutrient concentrations from each site. We believe this approach will provide readers the maximum amount of information and will make comparison to other work easier, since reduction of site data to medians is common in the literature (e.g., Robertson et al., 2001; Wang et al., 2007; Stevenson et al., 2012).

1.7. Concern that MT DEQ has Recommended Nutrient Criteria Concentrations in some Ecoregions Beyond the Applicable Reference Distribution

Two reviewers were concerned that nutrient criteria concentrations had been set at levels beyond any single observation collected in the regional reference streams. The Northern Rockies and Idaho Batholith are good examples. Although one reviewer agreed with MT DEQ that one should not use reference condition nutrient concentrations *alone* to set criteria, at the same time the idea of setting a criterion higher than the highest observation in the regional reference sites was clearly troubling to reviewers.

RESPONSE: The reviewers comments can be summarized as (1) concentrations beyond the reference distribution 75th percentile may be linked to known harm-to-use thresholds (e.g., via benthic algae density), but they will not be protective of sensitive, low-nutrient adapted organisms, and (2) in the ecoregions with naturally-low nutrients the harm-to-use concentrations derived from the dose response studies always had a range, and MT DEQ should have picked the lower concentration threshold given that we are operating beyond the bounds of the regional reference condition. Regarding point 1, we agree with the reviewer that if stream concentrations rise to the criteria and the criteria are beyond the reference condition, some organisms—like low-nutrient diatom taxa—would be displaced. The difficulty with establishing criteria to protect microscopic organisms like this is that there is no definitive harm to the beneficial uses established in Montana law. Studies generally show that with some additional nutrients ultra-oligotrophic streams will have more of the same macroivertebrates (as evidenced by O/E scores >1.0) and more robust populations of some fish. Fish—and to a somewhat lesser degree macroinvertebrates—link directly to Montana's beneficial uses. But as Montana state law is currently written, it would be difficult to defend a criterion based on protecting low-nutrient diatoms (as suggested by one reviewer).

Regarding the 2nd point, there is definitely merit to the idea that if there are several dose-response studies for an ecoregion and the concentrations from them generally fall beyond the reference distribution, greater weight should be given to the study or studies with the lower concentrations. As a result, in the final draft we have somewhat lowered the criteria recommendations in several ecoregions where this occurred. We have still kept an eye on maintaining the reference Redfield ratio, and in some cases the final concentrations are still beyond the reference distribution, but they are closer to it.

2.0 Selected Comments from Individual Reviewers

Here are important comments unique to individual reviewers.

2.2 How a Stream Reach is Delineated (Pertains to Assessment Methodology)

One reviewer was concerned that the flexible manner by which a stream reach can be delineated could make it difficult for any stream reach to ever be found impacted by nutrients, because data from impacted sites would be lumped with data from unimpacted sites and would, in effect, dilute the signal. The reviewer also noted that because of the flexibility in establishing assessment reaches, intentional manipulation of reach lengths could drive the outcome.

RESPONSE: The potential for unethical actions to manipulate analysis outcomes is always present in assessment work, but the high level of professionalism in the MT DEQ staff is such that this issue has not arisen. Regarding the flexibility of assessment reach lengths, this was done purposefully as discussed in detail in Appendix A.2.0 of Suplee and Sada de Suplee (2011). A basic assumption of the method is that reaches should be relatively homogenous in time (over the past 10 years) and in space, and observations collected within the reach should be largely independent. One reviewer was concerned about sample independence but MT DEQ has demonstrated independence in similar nutrient-concentration datasets using standard statistical tests (Durbin-Watson, Rank von Neumann). From these results and earlier experience, temporal and spatial independence guidelines were defined to make sure data collection maintains sample independence to the degree possible (nutrient and biological samples have to be collected a month apart at a site, for example).

If an assessor concludes that a reach is really *not* adequately homogenous (e.g., it comprises a distinctly impacted segment and an unimpacted segment[3]) it is incumbent upon the assessor to subdivide the reach and make an independent assessment of each new segment. This stratification allows maximal precision of estimates for minimal sampling effort (Norris et al., 1992). What remains constant is the minimum number of water quality and biological samples that need to be collected in <u>each</u> of these new assessment reaches in order to make a final compliance decision. The reviewer seemed to suggest that fixed reach lengths, numbers of sites, etc. along streams would be better, but experience has shown that this is highly impractical in applied assessment. If MT DEQ were to carryout assessments using fixed-length reaches, results would be far more arbitrary then the approach currently found in the SOP.

3.0 Summary of Changes Resulting from the Peer Review

1. We have included soluble nutrient data in the final report (Update 1) for the River Breaks.

2. The Brillouin Evenness Index formula is provided in Update 1 with a better explanation of why this approach was taken. We have also characterized reference using median datasets, further described below in 3.

3. Reference condition within an ecoregion has been characterized by <u>first</u> reducing data from each reference site to a site median, <u>then</u> calculating distribution statistics (25th, 50th, 75th, and 90th) for the ecoregion based on the population of site medians.

4. We agree that there is merit to the idea that if there are several dose-response studies for an ecoregion and the concentrations from them are generally beyond the reference distribution, greater weight should be given to the study or studies with the lower concentrations. As a result, in the final document we have somewhat lowered the criteria recommendations in several ecoregions where nutrient concentrations are naturally very low. We have still kept an eye on the Redfield ratio of the regional reference streams and, in some cases, the final criteria recommendations are still beyond the reference distribution, but they are closer to it.

4.0 References

Bahls, L., R. Bukantis, and S. Tralles, 1992. Benchmark Biology of Montana Reference Streams. Helena, MT: Montana Department of Health and Environmental Sciences.

Bothwell, M.L., 1989. Phosphorus-limited Growth Dynamics of Lotic Periphytic Diatom Communities: Areal Biomass and Cellular Growth Rate Responses. Canadian Journal of Fisheries and Aquatic Sciences 46: 1293-1301.

DEQ-7, 2012. Circular DEQ-7, Montana Numeric Water Quality Standards. October 2012. Helena MT: Montana Department of Environmental Quality.

---

[3] Temporal non-uniformity must also be considered. For example, if a new feedlot with several permit violations was built alongside the stream three years ago, the assessor would not be including data in the analysis collected six years ago.

Flynn, K., and M.W. Suplee, 2013. Using a Computer Water Quality Model to Derive Numeric Nutrient Criteria. Lower Yellowstone River, MT. WQPBDMSTECH-22. Helena, MT: Montana Department of Environmental Quality, 269 p plus appendices.

Horner, R.R., Welch, E.B., and R.B. Veenstra,1983. Periphyton of Freshwater Ecosytems. Wetzel,R.G. (ed.) Dr. W. Junk Publishers, The Hague.

Hurlbert, S.H., 1984. Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs 54: 187-211.

Mapstone, B. D., 1995. Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors. Ecological Applications 5: 401-410.

Norris, R. H., E. P. McElravy, and V. H. Resh, 1992. "The Sampling Problem," in *The River Handbook*, Calow, P. and Petts, G. E., (Oxford, England: Blackwell Scientific Publications)

Rier, S.T., and R. J. Stevenson, 2006. Response of Periphytic Algae to Gradients in Nitrogen and Phosphorus in Streamside Mesocosms. Hydobiologia 561: 131-147.

Robertson, D.M., D. A. Saad, and A.M. Wieben, 2001. An Alternative Regionalization Scheme for Defining Nutrient Criteria for Rivers and Streams. U.S. Geological Survey, Water Resources Investigations Report 01-4073.

Stevenson, R.J., B.J. Bennett, D. N. Jordan, and R.D. French, 2012. Phosphorus Regulates Stream Injury by Filamentous Green Algae, DO, and pH with Thresholds in Responses. Hydrobiologia 695: 25-42.

Suplee, M.W., A. Varghese and J. Cleland, 2007. Developing Nutrient Criteria for Streams: An Evaluation of the Frequency Distribution Method. Journal of the American Water Resources Association 43: 453-472.

Suplee, M.W. V. Watson, A. Varghese, and J. Cleland, 2008. Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers. Helena, MT: Montana Department of Environmental Quality.

Suplee, M.W., and R. Sada de Suplee, 2011. Assessment Methodology for Determining Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels. WQPBMASTR-01. Helena, MT: Montana Department of Environmental Quality. *Available at*: http://deq.mt.gov/wqinfo/qaprogram/sops.mcpx

Vitousek, P.M., J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson, D.W. Schindler, W.H. Schlesinger, and D.G. Tilman, 1997. Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. Ecological Applications 7: 737-750.

Wang, L., D.M. Robertson, and P.J. Garrison. 2007. Linkages Between Nutrients and Assemblages of Macroinvetebrates and Fish in Wadeable Streams: Implication to Nutrient Criteria Development. Environmental Management 39: 194-212.